# Remote Imaging Applied to Schistosomiasis Control: The Anning River Project

Edmund Y Seto (1),* Don R Maszle (1), Robert C Spear (1), Peng Gong (1), Byron Wood (2)
(1) University of California, Berkeley, CA; (2) NASA Ames Research Center, Moffett Field, CA

**Abstract**

Landsat Thematic Mapper (TM) imagery was used to identify habitat suitable for *Oncomelania hupensis*, the snail vector for schistosomiasis in the Anning River Valley in Sichuan, China. The location of 55 snail habitat sites and 48 non-habitat sites were determined by GPS measurements. Landsat TM data were found to be quite variable for both snail and non-snail sites. Because of this, supervised maximum likelihood classification produced poor accuracy. It was hypothesized that the variability was due to the existence of multiple microenvironments, each with distinct spectral properties and each suitable as snail habitat. A two-tiered classification approach was developed in which an unsupervised classification was first performed for the snail and non-snail habitat data to generate five snail and five non-snail clusters. The signatures of the 10 clusters were then used to perform maximum likelihood classification. Using this approach, 90.3% of the snail habitat and 86.6% of the non-habitat were correctly identified. These results suggest that remote sensing may be an effective tool for identifying the habitat of the schistosomiasis vector in China. If so, this provides a surveillance method for studying the area affected by the new Three Gorges Dam, where profound ecological change will occur and schistosomiasis is predicted to become a major problem.

Keywords: schistosomiasis, remote sensing, Landsat, China, *Oncomelania hupensis*

## Introduction

The use of satellite imaging to remotely detect areas of high risk for transmission of infectious disease is an appealing prospect for large-scale disease monitoring. The detection of large-scale environmental determinants of disease risk, often called landscape epidemiology, has been motivated by several authors (1,2). The basic notion is that large-scale factors such as population density, air temperature, hydrological conditions, soil type, and vegetation can determine in a coarse fashion the local conditions contributing to disease vector abundance and human contact with disease agents. These large-scale factors can often be remotely detected by sensors or cameras mounted on satellite or aircraft platforms and can thus be used in a predictive model to mark high-risk areas of transmission and to target control or monitoring efforts. A review of satellite technologies for this purpose was recently presented by Washino and Wood (3), Hay (4), and Hay et al. (5).

In China, there is currently concern about the establishment and spread of infectious diseases, including malaria and schistosomiasis, in the area along the Yangtze

* Edmund Y Seto, University of California-Berkeley, EHS, School of Public Health, 140 Warren Hall, University of California, Berkeley, CA 94720 USA; (p) 510-649-8152; E-mail: edmund@sparky.berkeley.edu

upstream of the Three Gorges Dam, which is now under construction. Our group has been working with parasitologists from the Sichuan Institute of Parasitic Disease (SIPD) responsible for schistosomiasis monitoring and control in the area of the dam. The profound ecological and social changes that will take place as the dam is being constructed and after its completion may create new habitat for the snail species central to the cycling of the disease, as well as new relationships between humans, domestic animals, and the aquatic environment. The size of the lake that will be created behind the dam and the difficulty of access to this mountainous area make remote sensing technology an attractive adjunct to land-based surveillance of these changes as the lake fills and the dam goes into operation.

To explore the possible use of remote sensing in schistosomiasis control prior to the completion of the Three Gorges Dam, we have been studying a region where the disease is endemic, where ground-based data sets on disease prevalence and snail habitat exist, and which is of a scale suitable for study using remote sensing. With the assistance of our colleagues in the SIPD, we have focused on the area along the Anning River in the Daliang mountainous area of southwestern Sichuan Province. This region includes villages studied in our earlier work.

Remote sensing has been demonstrated to be a viable means of identifying habitat for vectors of other diseases. The potential efficacy for using remote sensing to determine high-risk areas of malaria transmission was recently illustrated (6,7). Two types of Anopheline mosquito habitat—unmanaged pastures and transitional swamps—were shown to be detectable based on classification of Landsat Thematic Mapper (TM) data. That research was an extension of previous work that focused on the identification of high and low Anopheline-producing rice fields (8). Landsat TM data have also been used to map land cover to study landscape correlates of Lyme disease (9). In that study, disease data and landscape classifications were overlaid to look for land cover correlates to disease risk.

Several studies have implied that remote sensing could be a useful tool for schistosomiasis monitoring. Cross and Bailey (10) and Cross et al. (11) showed a correlation between local temperature variation and prevalence rate. Malone et al. (12) showed that historical prevalence data correlated well with remotely detectable geographic features. Both of these studies took a different approach from the Anopheline studies in that they demonstrated a correlation between disease and ecological factors, whereas the malaria vector studies by Beck et al. (6,7) and Wood et al. (8), use remote sensing to identify habitat correlated with the presence of the disease vector.

In the current study, we ask if the second approach is applicable to detecting spatial variations in the vector population that transmits the parasite causing schistosomiasis japonicum, the Asian form of schistosomiasis. The disease vector, or, more appropriately, the intermediate host for schistosomiasis japonicum, is an amphibious snail, *Oncomelania hupensis*. A recent preliminary study by the SIPD used Advanced Very High Resolution Radiometer (AVHRR) data to identify snail habitat (13). In the current analysis we use higher resolution Landsat TM data to look for correlations with detailed ground-based snail ecology surveys. If surveyed snail habitats correlate with the satellite data, there is the potential to use remote sensing to monitor large and remote areas in the region of the dam, and to identify areas at high risk of transmission.

The current problem is different from that of detecting malaria vectors. The vector habitat for *O. hupensis* is usually a microenvironment that is itself not detectable using most remote sensing data because of their coarse spatial resolution. However, microenvironmental conditions may be affected by larger-scale factors including local vegetation type and surrounding crops, fertilizer usage, and water and temperature patterns. These factors will cause local changes in the environment, which in turn will influence the remote sensing signal. Further, the other two schistosomiasis studies found correlations between large-scale phenomena and disease rates, implying that something can be seen at this scale. The question addressed at present is whether remote sensing data of local areas can be accurately classified, based on large-scale environmental factors, so as to identify habitats that are suitable for these vector snails, and thus at high risk for transmission.

## Methods

To address this issue, our group conducted a study in the Anning River Valley in southwestern Sichuan Province. The Anning River Valley is a high mountain valley at an elevation of about 1,500 meters (m). This is primarily an agricultural area with irrigated farming of rice, corn, wheat, a variety of vegetables, and some export crops. The valley is also a highly endemic area for schistosomiasis japonica. The remote sensing data used were from the Landsat TM sensor. The ground data indicating suitable snail habitat were point observations from one environment type and were classified as habitat or non-habitat. Suitability was determined by the presence or absence of young or reproducing snails. Few locations are found with only adult snails present, presumably because snails leave unsuitable locations or die.

A large-scale snail monitoring effort was conducted in 1994 by the Xichang County Anti-Endemic Station (XCAS). The station is responsible for monitoring and controlling human schistosomiasis infection and vector snail ecology in the 17-township middle section of the Anning River Valley. Snail surveys were performed throughout the area in townships where the human incidence exceeded 10%. Snail surveillance was done in June. We chose this section of the river valley as our study area to take advantage of these existing surveillance data. The study area extends from Lizhou Township in the north to Hexi Township in the south, and covers about 45 km of the river valley around Xichang City. A map of the area showing these reference points is shown in Figure 1.

Two Landsat TM scenes (one spring, April 7, 1994, and one fall, October 16, 1994) were obtained for the region. Both images were free of cloud cover over the area of interest, and each represents a distinct agricultural season. The major crops during these times are rice and corn in summer-fall and wheat and beans in the winter-spring season.

Ground data on the locations of snail colonies were obtained from the XCAS's 1994 snail surveys (this is being supplemented with density information). For 10 days in the middle of June 1997, our group, with the help of the local authorities and the head of the XCAS, visited townships and recorded the geographic locations of the 1994 surveillance data. Collection sites were located with a Trimble Pro XL global positioning system (GPS) to allow for correlation with the remote sensing data. Three base stations were established and positioned with respect to a known surveyed control point at the peak of the Lushan mountain southeast of Xichang City. All data points were differentially corrected to the base station locations to provide positioning accuracy in the 1 to 5 m range.

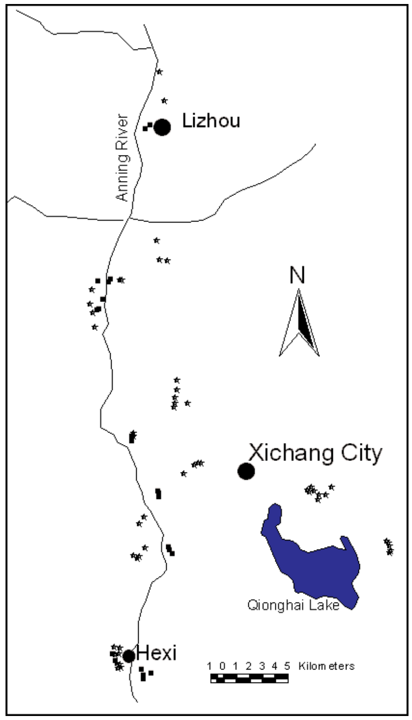


**Figure 1** Map of the Anning River Valley study area. Points represent snail habitat and non-habitat sites distributed in the valley from Lizhou Township in the north to Hexi Township in the south.

Collection sites were located in 14 townships throughout the study area. Townships were chosen based on availability of 1994 data or if there was historical knowledge of apparently stable snail habitat or non-habitat. Three environment types exist in the study area: irrigated farming in the river plain, terraced rice culture at the base of the hills, and mountain stream areas higher in the mountains. The three habitat types are structurally different with distinct local ecologies. In light of this, the study was limited to one type of environment, irrigated farming areas in the river plain, for which there was an abundance of ground/field data (and travel was more convenient). Snail habitat in the river plain area is limited to irrigation and drainage ditches and the boundaries of fields. This resulted in a total of 103 data points (55 classified as habitat and 48 as non-habitat).

Image processing was performed using PCIWORKS image processing software. Before data analysis, the images were geometrically corrected and registered using 11 ground control points taken throughout the river valley. Points used for referencing the image to a world coordinate system were large structures easily seen on the image, such as the corners of the Xichang airport runway, large intersections, and an isolated paved village compound. The 103 ground/field data points were located on the image. Each snail habitat and non-habitat site was specified as a 3- by 3-pixel area surrounding the site location as determined in the field by GPS measurements.

After geographic correction, a preliminary supervised maximum likelihood classification was performed using all TM channels from both dates. The 55 habitat and 48 non-habitat areas were used both to train the classification algorithm and assess the accuracy of the classification. The results of this accuracy assessment are presented in the next section.

Realizing that the accuracy of our preliminary classification was inadequate, we next employed a two-tiered analysis approach. The first step of this approach employed an unsupervised classification method called Isodata clustering to break up snail habitat and non-habitat classes into subclasses. The Isodata algorithm is an iterative process whereby the pixels of the image are grouped into clusters based on an examination of their multispectral brightness values. Pixels grouped into the same cluster have similar spectral properties. The Isodata algorithm was first applied to those pixels corresponding to snail habitat sites. The algorithm was used to categorize the pixels into five separate clusters. These five snail habitat clusters may correspond to different microhabitats that are all suitable for snails. The Isodata algorithm was then run using the non-habitat sites to produce five non-habitat clusters. The spectral distributions for each of these 10 clusters were determined and used to perform the second part (i.e., the supervised maximum likelihood classification) of this two-tiered analysis.

## Results

The results of the preliminary supervised classification using all TM bands from the spring and fall images are presented in Table 1. For the 55 snail habitat sites, there was good classification accuracy, with 89.3% of the pixels being classified correctly. However, for the non-habitat sites there were many misclassified pixels, with only 52.3% of the pixels being accurately classified as non-habitat. Among unclassified pixels, 3.4% of them corresponded to snail habitat sites and 8.8% of them corresponded to non-habitat sites.

**Table 1** Results of Preliminary Maximum Likelihood Classification of Snail Habitat and Non-Habitat Sites

| | Total # Pixels | % Unclassified Pixels | % Classified as Snail Habitat | % Classified as Non-Habitat |
|---|---|---|---|---|
| 48 Non-habitat sites | 432 | 8.8 | 38.9 | 52.3 |
| 55 Snail habitat sites | 495 | 3.4 | 89.3 | 7.3 |

Table 2 shows the result of the two-tiered classification. For the pixels corresponding to the 55 snail habitat sites, 3.6% were unclassified. Of the remaining 96.4%, 90.3% of the pixels were correctly classified as snail habitat. For the pixels corresponding to the 48 non-habitat sites, 4.2% were unclassified. Of the remaining 95.8%, 86.6% of the pixels were correctly classified as non-habitat. Table 3 presents a classification matrix showing the percentages of each cluster for both types of habitat.

The resulting classification for the Anning Valley is shown in Figure 2. A 5- by 5-pixel mode filter was applied to the image for presentation. The mode filter is primarily used to clean up thematic maps for presentation purposes by grouping together areas that are predominantly snail habitat or non-habitat. More specifically, for each 5- by 5-pixel area, the predominant class is assigned to all pixels in the area.

**Table 2** Results of Two-Tiered Analysis Using Isodata and Maximum Likelihood Classification Algorithms

| | Total # Pixels | % Unclassified Pixels | % Classified within Snail Habitat Clusters | % Classified within Non-Habitat Clusters |
|---|---|---|---|---|
| 48 Non-habitat sites | 432 | 4.2 | 12.6 | 83 |
| 55 Snail habitat sites | 495 | 3.6 | 87.1 | 9.2 |

**Table 3** Percentage of Pixels Classified by Cluster for Snail Habitat and Non-Habitat Sites

| | | | Snail Habitat Clusters | | | | | Non-Habitat Clusters | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total # Pixels | % Unclass-ified Pixels | % c1 | % c2 | % c3 | % c4 | % c5 | % c6 | % c7 | % c8 | % c9 | % c10 |
| 48 Non-habitat sites | 432 | 4.2 | 6.9 | 3.7 | 0.2 | 1.6 | 0.2 | 28.9 | 9.0 | 18.3 | 11.8 | 15.0 |
| 55 Snail habitat sites | 495 | 3.6 | 31.3 | 23.6 | 8.5 | 17.6 | 6.1 | 4.8 | 0.0 | 4.0 | 0.0 | 0.4 |

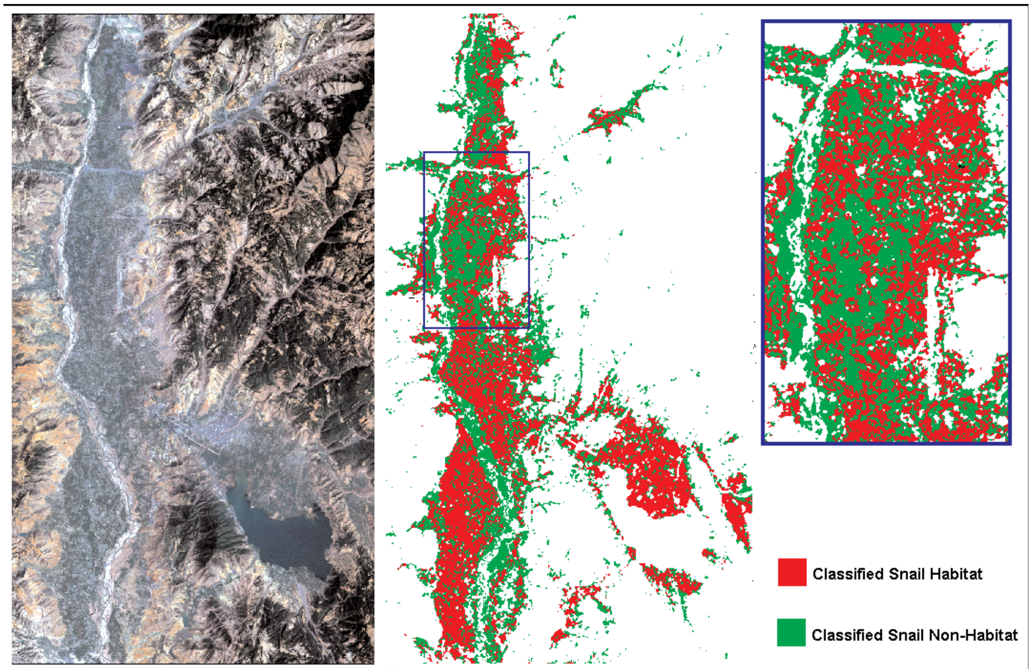


**Figure 2** Three panels showing (from left) (a) Landsat TM of Anning river valley, (b) classification of habitat using Isodata and maximum likelihood algorithms; (c) enlargement of valley floor showing mixed habitat.

## Discussion

Despite the fact that we limited our analysis to only those sites that were in the irrigated farming areas located in the river plain, there was a great deal of variability within the snail habitat and non-habitat sites. This was observed visually in the field as well as in the distributions of the spectral data. Our preliminary classifications ignored this variability by lumping all of the habitat sites together and all of the non-habitat sites together to train the classification. As a result, the snail habitat class included many of the non-snail sites, while the non-habitat class did not classify enough of the non-snail sites. This poor classification may be due to the existence within the irrigated farming environment of multiple microenvironments/habitats that each have distinct spectral properties. Hence, the terms "snail habitat" and "non-habitat" encompass distinctly different microenvironments that support or do not support snails, respectively. Therefore, when either snail habitat or non-habitat is considered as a whole, it appears to be quite variable.

In the two-tiered approach, we solved the problem of multiple microenvironments by using the Isodata algorithm to effectively separate the highly variable habitats into relatively 'pure,' less variable clusters before performing supervised classification. This was not based on field observation; rather, the spectral data were used to create these clusters. The choice to create five habitat clusters and five non-habitat clusters is not explained in detail because these numbers were chosen somewhat arbitrarily. The high classification accuracy, however, indicates that such numbers are not unreasonable. It will not be hard to fine-tune the number of clusters by looking at the variability and separability between signatures.

In addition to refining the number of clusters, we are also working on reducing the number of bands used to include only those that add information to the classification. Once we have reduced the classification down to the key bands, we hope to develop an understanding of what the clusters correspond to in the field.

## Future Work

Although our results thus far are quite promising, we acknowledge that they are still very much preliminary in nature. There is considerable work to be done before the methods described here can be applied to the field in the form of disease surveillance. One of our first aims is to develop a better sense of the accuracy—and, thus, limitations—of remote sensing classifications applied to schistosomiasis.

In our current work we assessed the accuracy of the classification only at the locations of the training sites. Using the same data for the training and validation of the classification may have resulted in artificially high accuracies. We plan to revisit the Anning River Valley to validate our two-tiered analysis with a rigorous field study. For this field study we first intend to obtain spring and fall Landsat TM images from a more recent year than 1994 to repeat the two-tiered classification. This more recent classification would then be validated in the field. We intend to sample pixels from this more recent classification and visit the corresponding locations in the field. True snail status would be assessed for each location and a better assessment of classification accuracy would be produced.

Another study we plan to conduct will assess the degree to which additional ground data might improve the classification accuracy. According to SIPD (14), the ecological correlates of *O. hupensis* snail habitat in Sichuan include the existence of certain vegetation types; size and density of irrigation ditches; proximity of agricultural field edges; wet lowland areas; soil moisture, type and quality; and, local temperature. In addition, it is our hypothesis that to some degree the chemical properties of the soil and water condition the existence of snails at a particular site. Some information, such as temperature, soil type, and vegetation type and coverage are available at a coarse scale for the Anning Valley, while other data, particularly the soil and water chemistry data, will have to be measured in the field during our randomized field validation study.

Several issues will have to be addressed when dealing with such multivariate data that are measured on several different scales and with varying reliabilities. For example, soil type data is a nominal variable and percent vegetation coverage, a bounded, interval variable. Because traditional remote sensing image analysis algorithms such as the maximum likelihood classifier cannot be used to process nominal and ordinal data, we will analyze these additional ground data using several non-traditional techniques: CART (15), logit regression (16), evidential reasoning (17,18), and artificial neural algorithms (19). Each of these algorithms can handle all the different levels of measurements and have proven useful in classification tasks where similar issues existed.

These multivariate approaches can be used to develop a classification for snail habitat based on ecology. The accuracy of this ecological classification can be compared with that of our remote sensing classification algorithm to gauge the added importance of incorporating ground ecology measurements in our classification of snail habitat. Furthermore, the ecological classification will help in developing an ecological interpretation of the remote sensing classification algorithm, which is central to being able to extrapolate the use of the algorithm to different areas and to different snail subspecies. Of particular interest is the determination of whether the distinct habitat clusters identified in our remote sensing classification correspond to distinct ecological conditions in our ecological classification, and later, how both the remote sensing and ecological classifications change between different schistosomiasis-prevalent regions in China.

The work described thus far has focused on locating snail habitat. Although the existence of snails is a necessary criterion for disease transmission, it does not serve as an accurate indication of disease prevalence since, within areas where snails exist, the extent of human and animal infection vary considerably. Moreover, in some locations where snails exist, no disease transmission occurs at all. It is clear, however, that on a local scale, infection intensity and disease prevalence are related to the relationships between people, animals, and snails, as they may be mediated by landscape features. Alongside our remote sensing work, we have been working with mathematical models as a way to better understand such site-specific factors at the local level and their impact on the dynamics of disease transmission.

From a remote sensing standpoint, however, many of the landscape features that are related to infection intensity—including the nature and density of irrigation in villages, and the proximity and density of settlements—can be identified and quantified using remote sensing technologies. In addition, topographical features such as slope and aspect determine the flow of water channels, which may influence the transmission of disease. Therefore, it is of considerable interest to determine if topographical or landscape features that can be determined remotely are correlates of transmission. Such information would further inform remote surveillance programs for prioritizing locations within the Three Gorges region for intensive ground investigation. To investigate these questions, higher resolution images than those from Landsat TM would be necessary. A future study will look at aerial photographs and/or higher resolution satellite images, such as those from SPOT HRV-PAN and IRA-1D, which are available now, and from Space Imaging and Earth Watch, which might become available soon.

## Acknowledgments

This project has been partially supported by a NASA-Ames Joint Research Interchange (NCC2-5102) and a University of California Pacific Rim Research Grant.

## References

1. Pavlovsky E. 1966. *The natural nidality of transmissible disease.* Urbana, IL: University of Illinois Press.
2. Meade M, Florin J, Gesler W. 1988. *Medical geography.* New York: The Guilford Press.
3. Washino R, Wood B. 1994. Application of remote sensing to arthropod vector surveillance and control. *American Journal of Tropical Medicine and Hygiene* 50(6):134–44.
4. Hay S. 1997. Remote sensing and disease control: Past, present and future. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 91:105–6.
5. Hay S, Parker M, Rogers D. 1997. The impact of remote sensing on the study and control of invertebrate intermediate hosts and vectors of disease. *International Journal of Remote Sensing* 18(14):2899–2930.
6. Beck L, Rodriguez S, Dister A, Rodriguez A, Washino R, Roberts D, Spanner M. 1997. Assessment of a remote sensing based model for predicting malaria transmission risk in villages of Chiapas, Mexico. *American Journal of Tropical Medicine and Hygiene* 56(1):99–106.
7. Beck L, Rodriguez M, Dister S, Rodriguez A, Rejmankova E, Ulloa A, Meza RA, Roberts D, Paris J, Spanner M, Washino R, Hacker C, Legters L. 1994. Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. *American Journal of Tropical Medicine and Hygiene* 51(3):271–80.
8. Wood B, Washino R, Beck L, Hibbard K, Pitcairn M, Roberts D, Rejmankova E, Paris J, Hacker C, Salute J, Sebesta P, Legters L. 1991. Distinguishing high and low anopheline-producing rice fields using remote sensing and GIS technologies. *Preventive Medicine* 11:277–88.
9. Dister S, Beck L, Wood B, Falco R, Fish D. 1993. *The use of GIS and remote sensing technologies in a landscape approach to the study of Lyme disease transmission risk.* GIS '93 Symposium, Vancouver, BC.
10. Cross ER, Bailey RC. 1984. Prediction of areas endemic for schistosomiasis through use of discriminant analysis of environmental data. *Military Medicine* 149(1):28–30.
11. Cross ER, Sheffield C, Perrine R, Pazzaglia G. 1984. Predicting areas endemic for schistosomiasis using weather variables and a Landsat data base. *Military Medicine* 149(10):542–4.
12. Malone J, Huh O, Fehler D, Wilson P, Wilensky D, Holmes R, Elmagdoub A. 1994. Temperature data from satellite imagery and the distribution of schistosomiasis in Egypt. *American Journal of Tropical Medicine and Hygiene* 50(6):714–22.
13. Li Z, Yuan P, Yin R, He S, Gu X, Zhao W, Xu F. 1990. Identification of distribution area of *Oncomelania* by remote sensing technique. *Acta Scientiae Circumstantiae* 10(2).

14. Gu X. 1995. *Report of national key project of the 8th five-year plan.* Chengdu, China, Sichuan Institute of Parasitic Disease.

15. Breiman L, Friedman J, Olshen R, Stone C. 1984. *Classification and regression trees.* Monterey: Wadsworth and Brooks/Cole.

16. Chung C, Moon W. 1991. Combination rules of spatial geoscience data for mineral exploration. *Geoinformatics* 2:159–69.

17. Wang Y, Civco D. 1994. Evidential reasoning-based classification of multisource spatial data for improved land cover mapping. *Canadian Journal of Remote Sensing* 20(4):381–95.

18. Gong P. 1996. Integrated analysis of spatial data from multiple sources: using evidential reasoning and an artificial neural network for geological mapping. *Photogrammetric Engineering and Remote Sensing* 62(5):513–23.

19. Gong P, Pu R, Chen J. 1996. Mapping ecological land systems and classification uncertainty from digital elevation and forest cover data using neural networks. *Photogrammetric Engineering and Remote Sensing* 62(11):1249–60.